Number of Loops of Size h in Growing Scale-Free Networks

Ginestra Bianconi and Andrea Capocci
Département de Physique, Université de Fribourg, Pérolles, CH-1700, Switzerland
(Received 17 November 2002; published 21 February 2003)

The hierarchical structure of scale-free networks has been investigated focusing on the scaling of the number $N_h(t)$ of loops of size h as a function of the system size. In particular, we have found the analytic expression for the scaling of $N_h(t)$ in the Barabási-Albert (BA) scale-free network. We have performed numerical simulations on the scaling law for $N_h(t)$ in the BA network and in other growing scale-free networks, such as the bosonic network and the aging nodes network. We show that in the bosonic network and in the aging node network the phase transitions in the topology of the network are accompained by a change in the scaling of the number of loops with the system size.

DOI: 10.1103/PhysRevLett.90.078701 PACS numbers: 89.75.Hc, 89.75.Da, 89.75.Fb

Natural and social systems ranging from protein-protein interactions [1] to the World Wide Web [2] can be represented as *scale-free* (SF) networks [3–6], that is, sets of nodes connected by links with special statistical properties. These systems are characterized by large fluctuations in the individual *degrees*, i.e., in the number of links pointing or leaving each node. In a SF network, the statistical distribution $P(k)$ of the degree k is a power law $P(k) \sim k^{-\gamma}$. Accordingly, if $\gamma < 3$, the individual degree distribution displays infinite variance in the limit of infinite network size.

Besides this striking property, models traditionally defined on regular lattices display peculiar effects when defined on a scale-free topology instead [7–12]. Such phenomena have increased the interest in finding universal mechanisms to generate SF networks, which seem to arise independently in so diverse contexts.

Recently, these properties have been reproduced by the Barabási-Albert (BA) dynamical model of a random network. Its simple and natural algorithm, based on a "preferential attachment" rule [2,13], triggered an avalanche of research activity in the field and generated a rich "zoology" of SF network models sharing the same fundamental ingredients. Though recently alternative mechanisms [14,15] for the generation of scale-free networks have been proposed, in this work we restrict our attention to growing scale-free networks with preferential attachment. But the degree distribution is not the only relevant topological quantity. For the characterization of networks it is necessary to look at the motifs [16–18] recurrent on them and, in particular, at the number of loops of length larger than 3 [19]. In this work we derive a general formula for the number of loops of size h that generalizes the known result concerning triangular loops [20] ($h = 3$) for a BA network. In particular, we find that the number of loops $N_h(t)$ of size h at time t in a BA network scales as $N_h(t) \sim [m/2 \log(t)]^{\psi}$. In addition, we show that our formula is robust when tested on other growing SF network models, such as the bosonic network (BN) model [21] and the aging nodes (AN) model [22]. In the bosonic network the scaling of the number of loops with the system size changes below the Bose-Einstein phase transition, where a power-law scaling of the type $N_h(t) \sim t^{\xi}$ takes place.

The BA model [2,13] was the first and simplest algorithm generating scale-free undirected networks. In this model, a new node is added to the network at each time step, and it is connected by a fixed number of links m to highly connected existing nodes (preferential attachment). According to this rule, the probability Π_i that an existing node i at time t acquires a new link is assumed to be proportional to its degree $k_i(t)$. This reads

$$\Pi_i = m \frac{k_i(t)}{\sum_j k_j(t)}. \tag{1}$$

The model can be analyzed by a mean field approximation [2]. By this approach, one finds that the average degree of a node i that entered the network at time t_i increases with time as a power law

$$k_i(t) = \frac{m}{2}\sqrt{\frac{t}{t_i}}. \tag{2}$$

A network built in this way displays a power-law degree distribution $P(k) \sim k^{-3}$. In order to investigate the inhomogeneous topology of the network, we define a loop of size h (an h loop) as a closed path of h links that visits each intermediate node only once.

In the case $m > 2$, the BA scale-free network is a very compact network, with loops of any size. As the network evolves, new loops are introduced in the network. By definition, new loops include the newly added node: indeed, a new h loop is formed if the new node is connected to two nodes already connected by a self-avoiding path of size $h - 2$. We indicate with $p_{i,k}$ the probability that the nodes i, k, attached to the network at time t_i, t_k, are connected by a link. The rate at which new loops of length h are formed is given by the probability that the new node k is linked to two existing nodes i and j times the probability $P_{i,j}^{h-2}(t)$ that they are already connected by a self-avoiding path of size $h-2$. Therefore, we write the following rate equation for the average number of h loops $N_h(t)$:

$$\frac{\partial\langle N_h(t)\rangle}{\partial t} = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} p_{i,k}p_{j,k}P_{ij}^{h-2}(t), \tag{3}$$

where the factor $\frac{1}{2}$ takes into account that each pair of nodes i, j has been counted twice in the sums.

On the other hand, two directly connected nodes i and j belong to an h loop if they are also connected by a self-avoiding path of length $h-1$. Let $P_{i,j}^{h-1}(t)$ be the probability that this path exists; the total number of h loops passing through node i is given by $\sum_{j=1}^{N} p_{i,j}P_{i,j}^{h-1}(t)$. We obtain the average number $\langle N_h(t)\rangle$ of h loops in the system times $2h$ by summing this quantity over all the nodes i in the network. In fact, each loop has been counted $2h$ times, because there are h nodes in the loop, and two possible directions. Hence, we can write

$$\langle N_h(t)\rangle = \frac{1}{2h}\sum_{i=1}^{N}\sum_{j=1}^{N} p_{i,j}P_{ij}^{h-1}(t). \tag{4}$$

By replacing Eq. (2) in Eq. (1), the probability that the node i is attached to node k is

$$p_{i,k} = \frac{m}{2}(1/\sqrt{t_i t_k}). \tag{5}$$

Moreover, the probability that a node k is connected with nodes i and j is proportional to the probability that i and j are already connected, i.e.,

$$p_{i,k}p_{j,k} = m\frac{1}{2t_k}p_{i,j}. \tag{6}$$

By replacing this result in (3) and by the definition (4), we obtain

$$\frac{\partial\langle N_h(t)\rangle}{\partial t} = \frac{m}{2t}(h-1)\langle N_{h-1}(t)\rangle. \tag{7}$$

Consequently, the rate at which new loops of size h are introduced in the system is proportional to the mean number of loops of size $h-1$.

Equation (7) has a recursive structure that allows its integration without any detailed information about the probabilities $P_{i,j}^{h}(t)$. In fact, the rate at which new loops of length h are formed can be expressed only in terms of the number of loops of minimal size (i.e., $h=3$),

$$\frac{\partial^{h-3}\langle N_h(\zeta)\rangle}{\partial\zeta^{h-3}} = (h-1)!\langle N_3(\zeta)\rangle \tag{8}$$

with $\zeta = \frac{m}{2}\log(t)$. The number of triangular loops $\langle N_3(\zeta)\rangle$ can be computed directly, for the triangular loops increase in time following (3),

$$\frac{\partial\langle N_3(t)\rangle}{\partial t} = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} p_{i,k}p_{j,k}p_{i,j}. \tag{9}$$

Since $p_{i,j}$ is given by Eq. (5), approximating sums by integrals we write the rate equation in the form

$$\frac{\partial\langle N_3(t)\rangle}{\partial t} = \frac{1}{2}\left(\frac{m}{2}\right)^3\int_0^t dt_i\int_0^t dt_j\frac{1}{t_i}\frac{1}{t_j}\frac{1}{t} = \frac{1}{2}\left(\frac{m}{2}\right)^3\frac{1}{t}[\log(t)]^2. \tag{10}$$

Integrating (10) we find, in agreement with [20],

$$\langle N_3(t)\rangle = \frac{1}{3!}\left[\frac{m}{2}\log(t)\right]^3. \tag{11}$$

Using Eq. (11) in Eq. (8), we compute the number of loops of size h, $\langle N_h(t)\rangle$ and we find

$$\langle N_h(t)\rangle = \left[\frac{m}{2}\log(t)\right]^h[1+O(\zeta^{-1})] \sim \left[\frac{m}{2}\log(t)\right]^h. \tag{12}$$

The expression for the scaling of $N_h(t)$ with the system size t in a BA network does not suggest a practical way to measure $N_h(t)$. To this purpose, one has to study the symmetrical adjacency matrix a of the network, whose generic element a_{ij} is defined by $a_{ij}=1$ if i and j are connected and $a_{ij}=0$ otherwise. Knowing this matrix, one directly measures the number of paths starting from a node i and returning on it after h steps that visit intermediate nodes only once. According to this argument, the term $N_h(t)$ has a dominating term of the type $\sum_i(a^h)_{i,i}/(2h)$ and subdominant terms excluding all trivial contributions coming from paths intersecting on themselves. Let us assume that the network does not contain self-loops; i.e., $a_{ii}=0$ for all i of the network. In this case, for $h=3$ we simply have

$$N_3 = \frac{1}{6}\sum_i(a^3)_{ii}. \tag{13}$$

For $h=4$ and $h=5$, by simple arguments it is possible to show that

$$N_4 = \frac{1}{8}\left[\sum_i(a^4)_{ii} - 2\sum_i(a^2)_{ii}(a^2)_{ii} + \sum_i(a^2)_{ii}\right] \tag{14}$$

and that

$$N_5 = \frac{1}{10}\left[\sum_i(a^5)_{ii} - 5\sum_i(a^2)_{ii}(a^3)_{ii} + 5\sum_i(a^3)_{ii}\right]. \tag{15}$$

Using relations (13)–(15), we can directly measure $N_h(t)$ for $h=3,4,5$ in the BA scale-free network model to check our analytical results.

In Fig. 1 we show that the scaling of $N_h(t)$ for a BA network with $m=2$ follows

$$N_h(t) \sim \left(\frac{m}{2}\log(t)\right)^{\psi(h)}. \tag{16}$$

The measured effective exponent $\psi(h)$ reported in the inset of Fig. 1 grows with h as expected. Nevertheless, $\psi(h)$ differs from the predicted asymptotic behavior $\psi(h)=h$ in a significant way. This can be explained

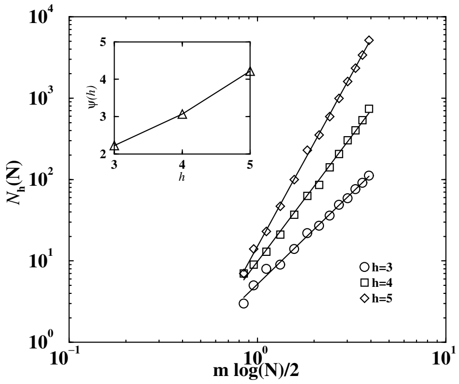

FIG. 1. Scaling of $N_h(t)$ as a function in a BA network for t up to $t = 10^4$. The inset shows the values of the measured exponent ψ defined in (16) for $h = 3, 4, 5$.

by the fact that we are considering networks of size up to $t = 10^4$ nodes and we are still far from the asymptotic behavior $t \to \infty$.

To check the robustness of the scaling relations of $N_h(t)$, we have measured the number of loops of size $h = 3, 4, 5$ in two alternative growing SF network models: the bosonic network [21] and the aging nodes network [22].

In the bosonic network, each node i is assigned an innate quality, represented by a random "energy" ϵ_i drawn from the probability distribution $p(\epsilon_i)$. The attractiveness of each node i is then determined jointly by its connectivity k_i and its energy ϵ_i. In particular, the probability that node i acquires a link at time t is given by

$$\Pi_i = \frac{e^{-\beta\epsilon_i} k_i(t)}{\sum_j e^{-\beta\epsilon_j} k_j(t)}; \tag{17}$$

i.e., low energy, high degree nodes are more likely to acquire new links. The parameter $\beta = 1/T$ in Π_i tunes the relevance of the quality with respect to the degree in the acquisition probability of new links. Indeed, for $T \to \infty$ the probability Π_i does not depend anymore on the energy ϵ_i and the BN model reduces to the BA model. On the other hand, in the limit $T \to 0$ only the lowest energy node has nonzero probability to acquire new links. Moreover, in Ref. [21] it has been shown that the connectivity distribution in this network model can be mapped on the occupation number in a Bose gas. According to this analogy, one would expect a corresponding phase transition in the topology of the network at some temperature value T_c.

In fact, for energy distributions such that $[p(\epsilon) \to 0$ for $\epsilon \to 0]$, one observes a critical temperature T_c. For $T > T_c$ the system is in the "fit-get-rich" (FGR) phase, where nodes with lower energy acquire links at a higher rate than higher energy nodes, while for $T < T_c$ a "Bose-Einstein condensate" (BE) or "winner-takes-all" phase emerges, where a single node grabs a finite fraction of all the links. We simulated this model assuming

$$p(\epsilon) = (\theta + 1)\epsilon^\theta \quad \text{and} \quad \epsilon \in (0, 1), \tag{18}$$

where $\theta = 1$. For this distribution, it has been shown that $T_c(\theta = 1) = 0.7$ [21]. In particular, in Fig. 2 we report $N_h(t)$ for $h = 3, 4, 5$ as a function of the number of nodes t in the network, at temperature $T = 1.5$ and $T = 0.5$, respectively.

According to simulations, in correspondence with the phase transition, the scaling of $N_h(t)$ drastically switches from a BA-like behavior

$$N_h(t) \propto \left(\frac{m}{2}\log(t)\right)^{\psi(h)} \quad \text{for } T > T_c \tag{19}$$

to a power-law scaling

$$N_h(t) \propto t^{\xi(h)} \quad \text{for } T < T_c. \tag{20}$$

We have measured the scaling of $N_h(t)$ with the system size t also for a growing network with aging nodes introduced in [22]. The model has been motivated by the observation that in many real networks, e.g., the scientific citations network, old nodes are less cited than recent ones. In the goal of representing this feature, the probability Π_i to attach a link to a node i arrived in the network at time t_i is modified to be

$$\Pi_i = \frac{(t - t_i)^{-\alpha} k_i(t)}{\sum_j (t - t_j)^{-\alpha} k_j(t)}, \tag{21}$$

where α is an external parameter. As in the BA model, a new node is connected to m existing nodes. The resulting structure of such a network strongly depends on the constant α. For $\alpha < 1$, the degree distribution is a power-law $P(k) \sim k^{-\gamma}$ with an exponent monotonically increasing from $\gamma = 2$ in the limit $\alpha \to -\infty$ to $\gamma \to \infty$ in the limit $\alpha \to 1$: on the other hand, for $\alpha > 1$ no power law is observed in the degree distribution. Therefore, this model reproduces a SF network only in the region $\alpha > 1$.

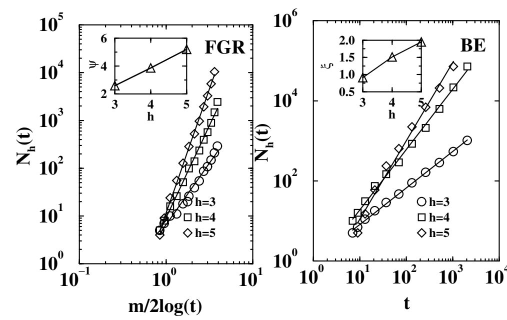

FIG. 2. Scaling of the number of loops $N_h(t)$ of size h with the system size t in the FGR and in the BE phase of a bosonic network. The plotted data have been obtained for a network with $p(\epsilon) = 2\epsilon$ and $\epsilon \in (0, 1)$ (which has $T_c = 0.7$ [19]) at $T = 1.5$ (FGR phase) and at $T = 0.5$ (BE phase). In the insets we report the value of the exponents ψ (2.6, 3.9, 5.2) and ξ (0.78, 1.28, 156) found in the simulations for $h = 3, 4, 5$, respectively.

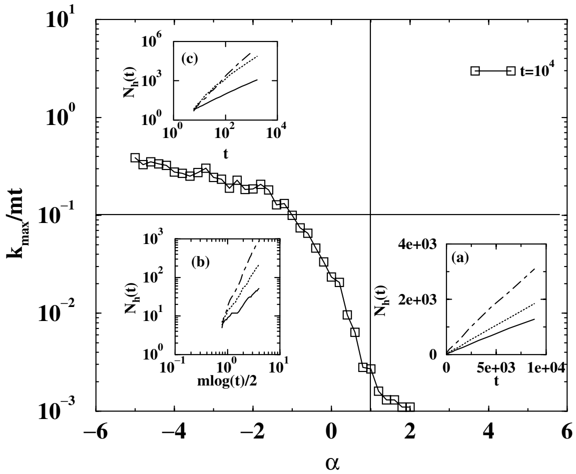

FIG. 3. The order parameter $k_{\max}/mt$ for a network with aging of the nodes, size $t = 10^4$ and $m = 2$. We distinguish among three regions of the phase space: $\alpha > 1$, $\alpha \in (-1, 1)$, and $\alpha < -1$. In the insets we report the typical behavior of $N_h(t)$ as a function of t for $h = 3, 4, 5$ in the three regions.

Moreover, as observed in [22], in the limit $\alpha \to -\infty$ the oldest node is connected to an increasing fraction of all the links, reminding the condensation observed in a bosonic network. To take into account this phenomenon, we introduce a value α^* such that for $\alpha < \alpha^*$ the fraction of links attached to the most connected node exceeds a finite threshold F. We expect the scaling of $N_h(t)$ to be different in the three regions $\alpha > 1$, $\alpha \in (\alpha^*, 1)$, and $\alpha < \alpha^*$.

We measured the total fraction of links $k_{\max}/mt$ attached to the oldest node in a network with $m = 2$ and $t = 10^4$ nodes. The threshold has been fixed at $F = 0.1$, in order to distinguish the "condensate" phase from the simple scale-free phase. The value for α^* was found to be $\alpha^* = -1$. We then measured the number of loops of size $h = 3, 4, 5$ for networks made of up to 10^4 nodes in the three ranges of value of α. We have observed that, for $\alpha > 1$, the number of loops of size h scales linearly with t (at least for $h = 3, 4, 5$). In inset (a) of Fig. 3 we report the data for $\alpha = 1.5$. On the contrary, for $\alpha \in (-\alpha^*, 1)$ we measured the scaling

$$N_h(t) \propto \left(\frac{m}{2}\log(t)\right)^{\psi(h)} \qquad (22)$$

with $\psi(h)$ a monotonic function of h. In inset (b) of Fig. 3, data for $\alpha = 0.5$ are reported. Finally, in the region $\alpha < \alpha^* = -1$, $N_h(t)$ becomes proportional to a power law of the system size, as is shown in inset (c) of Fig. 3, referring to the case $\alpha = -5$.

In conclusion, we have introduced the number of h loops $N_h(t)$ for a network of t nodes as a characterizing quantity for random networks. We have observed that in the BA scale-free networks $N_h(t)$ scales as a power law of the logarithm of the system size. Moreover, we have observed that indeed this scaling seems to be a marking feature of growing scale-free networks with preferential attachment. In particular, topological phase transitions in the bosonic network and in the network including aging nodes are accompanied by a drastic change in the scaling of $N_h(t)$ with the system size.

The authors are grateful to Yi-Cheng Zhang and Alessandro Vespignani for their useful comments. The research presented in this Letter has been financially supported by the Swiss National Foundation, under Grant No. 2051-067733.02/1, and by the European Commission, Fet Open Project No. COSIN IST-2001-33555.

[1] H. Jeong, S. Mason, A.-L. Barabasi, and Z. N. Oltvai, Nature (London) **411**, 41 (2001).
[2] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).
[3] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[4] S. Strogatz, Nature (London) **410**, 268 (2001).
[5] S. N. Dorogovtsev and J. F. F. Mendes, Adv. Phys. **51**, 1079 (2002).
[6] M. E. J. Newman, in *Handbook of Graphs and Networks*, edited by S. Bornholdt and H. G. Schuster (Wiley-VCH, Berlin, 2002).
[7] R. Cohen, K. Erez, D. ben-Avraham, and S. Havlin, Phys. Rev. Lett. **85**, 4626 (2000); **86**, 3682 (2001).
[8] R. Pastor-Satorras and A. Vespignani, Phys. Rev. Lett. **86**, 3200 (2001).
[9] A. Aleksiejuk, J. A. Holyst, and D. Stauffer, Physica (Amsterdam) **310A**, 260 (2002).
[10] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, Phys. Rev. E **66**, 016104 (2002).
[11] M. Leone, A. Vazquez, A. Vespignani, and R. Zecchina, Eur. Phys. J. B **28**, 191 (2002).
[12] G. Bianconi, Phys. Lett. A **303**, 166 (2002).
[13] A.-L. Barabàsi, R. Albert, and H. Jeong, Physica (Amsterdam) **272A**, 173 (1999).
[14] G. Caldarelli, A. Capocci, P. De Los Rios, and M. A. Muñoz, cond-mat/0207366.
[15] G. Mukherjee and S. S. Manna, cond-mat/0207476.
[16] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, Science **298**, 824 (2002).
[17] K. Klemm and V. Eguíluz, Phys. Rev. E **65**, 057102 (2002); Science **298**, 824 (2002).
[18] G. Szabó, M. Alava, and János Kertész, cond-mat/0208551.
[19] G. Caldarelli, R. Pastor-Satorras, and A. Vespignani, cond-mat/0212026.
[20] P. M. Gleiss, P. F. Stadler, A. Wagner, and D. A. Fell, Adv. Complex Syst. **4**, 207 (2001).
[21] G. Bianconi and A.-L. Barabási, Phys. Rev. Lett. **86**, 5632 (2001).
[22] S. N. Dorogovtsev and J. F. F. Mendes, Phys. Rev. E **62**, 1842 (2000); **63**, 056125 (2001).